Economic Modelling 54 (2016) 26–39

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/ecmod

Nowcasting Czech GDP in real time☆

Marek Rusnák

Czech National Bank and Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic

ARTICLE INFO

Article history:
Accepted 2 December 2015
Available online xxxx

Keywords:
Dynamic factor model
GDP
Nowcasting
Real-time data

ABSTRACT

In this paper, we employ a Dynamic Factor Model (DFM) to nowcast Czech GDP. Using multiple vintages of historical data and taking into account the publication lags of various monthly indicators, we evaluate the real-time performance of the DFM over the 2005–2012 period. The main result of this paper is that the accuracy of model-based nowcasts is comparable to that of the nowcasts of the Czech National Bank (CNB). Moreover, combining the DFM and the CNB nowcasts results in more accurate performance than in the case of the individual nowcasts alone. Our results also suggest that foreign variables are crucial for the accuracy of the model, while omitting financial and confidence indicators does not worsen the nowcasting performance.

1. Introduction

Because of considerable publication delays in the release of GDP data, the current state of the economy is subject to sizeable uncertainty. Accurate and timely estimates of the current state of the economy are therefore especially important for policymakers, who make their decisions in real time. In the present turbulent situation, obtaining the most up-to-date forecasts of GDP, possibly after each new data announcement, is becoming even more important, for example, in the event of irregular monetary policy meetings in the midst of a crisis or other unexpected developments in the economy. Such up-to-date forecasts of Czech GDP produced in real time are the objective of this study.

Forecasters face several problems when producing predictions in real time. Macroeconomic variables are announced in a non-synchronous manner, that is, they have different publication lags. As a result, forecasters have to work with datasets that contain many missing observations towards the end of the sample (the so-called ragged end problem). Another problem forecasters typically face is the fact that data are sampled at different frequencies. Most of the traditional forecasting models — such as leading indicator models and classical vector autoregressions — cannot easily deal with these issues: they cannot utilize the most up-to-date data releases in a model-consistent fashion.

The nowcasting framework of Giannone et al. (2008) has become the workhorse model of short-term forecasters at many central banks and other institutions (for an extensive list of references see Bańbura et al., 2013). The framework is based on a dynamic factor model cast in the state-space representation and on the application of the Kalman filter to deal with mixed frequencies and unbalanced datasets.[1] The framework can accommodate a potentially large number of variables by summarizing the information with a few common factors, thus overcoming the so-called curse of dimensionality (Stock and Watson, 2002b; Bernanke and Boivin, 2003). An additional advantage of the framework is that it allows forecasters not only to predict variables of interest in real time, but also to interpret and comment on the sources of the changes in the forecasts. This provides a story-telling dimension and a deeper understanding of the forecast that is almost as important to policymakers as the accuracy of the forecast itself. This feature is missing from most of the statistical models that are currently used for near-term projections.

An additional challenge for real-time forecasters is the presence of data revisions. Typically, the forecasting exercises and model selection are performed using revised data. It is well known, however, that the revisions to macroeconomic data are frequent and large (Faust et al., 2005; Garratt and Vahey, 2006; Aruoba, 2008; Croushore, 2011; Fernandez et al., 2011). Therefore, working with the last available data may provide starkly different results than those obtained using real-time data (as documented by many studies: Robertson and Tallman, 1998; Faust et al., 2003; Orphanides, 2001; Kugler et al., 2005; Molodtsova et al., 2008; Marcellino and Musso, 2011; Ince and Papell, 2013). As for the properties of revisions to Czech GDP, in our previous research (Rusnak, 2013), we find that the revisions are relatively

☆ We would like to thank Michal Franta, Tomáš Havránek, Roman Horváth, Gabriel Perez Quiros, Christian Schumacher, and Luboš Růžička for their useful comments and discussions, and Marta Bańbura for sharing her Matlab code. We would also like to thank the Editor of the *Economic Modelling* and three anonymous referees for their comments and suggestions that helped to improve the paper. We acknowledge the support of the Czech National Bank, the Grant Agency of Charles University (grant # 888413), and the Grant Agency of Czech Republic (grant P402/12/G097). The opinions expressed in this paper are only those of the authors and do not necessarily reflect the position and views of the Czech National Bank or any other institution with which the authors are associated.

E-mail address: marek.rusnak@cnb.cz.

[1] Previous seminal contributions include Wallis (1986) and Evans (2005). See also Foroni and Marcellino (2013), who provide a survey of state-of-the-art mixed frequency models that can deal with ragged end problems.

http://dx.doi.org/10.1016/j.econmod.2015.12.010
0264-9993/

large. Performing a proper real-time forecasting exercise using Czech data therefore seems to be greatly needed.

The short-term forecasting performance of various models of Czech GDP has been assessed before by many studies (Benda and Ruzicka, 2007; Arnostova et al., 2011; Havranek et al., 2012; Horvath, 2012). Unfortunately, most of these studies do not account for publication lags and data revisions, which renders the relevance of their results to policymakers rather questionable.[2] Consider, for example, the official comments that the CNB makes after each release of GDP. Out of 32 comments published by the Czech National Bank (CNB) during the 2005–2012 period, roughly 17 of them mention revisions to the national accounts as one of the sources of the deviation of the official CNB forecasts from the announced data. Obviously, revisions must be therefore considered an important issue to policymakers. Truly real-time exercises to evaluate the performance of dynamic factor models in the presence of data revisions are still relatively scarce. The exceptions are Schumacher and Breitung (2008) for Germany, Camacho and Perez-Quiros (2010) for the euro area, and Bańbura et al. (2013) and Lahiri and Monokroussos (2013) for the US. To the best of our knowledge, we are the first to investigate the performance of forecasts of Czech GDP in a truly real-time setting that employs unrevised vintages of historical data.

In this paper, we focus on the performance of the DFM in obtaining accurate forecasts of the current quarter GDP growth (so-called nowcasts). Accurate nowcasts are important since they serve as inputs to the structural models that are used for medium to long-term prediction (the CNB uses a G3 DSGE model, see Andrle et al., 2009, for details). Furthermore, the CNB comments on the releases of the latest GDP growth figures and discusses the deviations from its official predictions. This makes the accuracy of CNB nowcasts of crucial importance.

Formal model-based forecasts are typically compared to naive benchmarks or to other competing models. Comparisons with official central bank forecasts are rare, but do exist, especially in the context of model combinations (Lees et al., 2007; Adolfson et al., 2007; Groen et al., 2009; Edge et al., 2010; McDonald and Thorsrud, 2011). A common finding of these studies is that the accuracy of model-based forecasts of GDP is comparable to that of the official forecasts of the respective central banks.[3] In this paper, we contribute to this literature by evaluating the performance of the dynamic factor model using Czech real-time data and comparing it with the accuracy of the nowcasts of the Czech National Bank.

Finally, we show how one can use the methodology of Bańbura and Modugno (2010) to decompose the updates of Czech GDP nowcasts into the contributions of the individual variables — so-called *news*. This is possible since the dynamic factor model produces forecasts for all of the variables included. One can then interpret changes in the forecasts stemming from the differences between the actual data released and their predicted values. For example, it is reasonable to assume that a higher-than-expected value of industrial production will cause the forecast to be revised upwards. The dynamic factor model can quantify such statements. Similar decompositions of forecast updates are now regularly used by many central banks (see for example ECB, 2008; Bundesbank, 2009) to enhance the understanding of their short-term forecasts.

Our results suggest that the nowcasting performance of the medium-scale DFM is comparable to the nowcasts of the Czech National Bank. In addition, we find that the simple average of the DFM and CNB nowcasts is more accurate than the nowcasts of the DFM and CNB alone. We also find that the DFM nowcasts add value to the CNB nowcasts: conditional on the CNB nowcast, on average, GDP growth turns out to be higher when the DFM nowcast is higher. Similarly to D'Agostino and Giannone (2012) we find that the relative performance of the DFM is better at times of crisis, which are characterized by large comovements of variables. We also find that the inclusion of foreign variables is crucial: if we exclude foreign variables the performance worsens significantly, while the omission of financial variables or surveys does not result in a dramatic deterioration of the forecasting accuracy.

The remainder of this paper is organized as follows. Section 2 briefly discusses the dynamic factor model, Section 3 describes our real-time dataset and provides details of the empirical exercise together with its results. Section 4 presents examples of nowcast update decompositions, while Section 5 provides further results and sensitivity checks. Section 6 concludes.

2. Dynamic factor model

Dynamic factor models aim at capturing the most important features of the data while remaining parsimoniously specified. They do so by assuming that the bulk of the comovements in macroeconomic variables are driven by just a few common factors (this seems to be the case in the US, see Giannone et al., 2005). The technology of dynamic factor models has evolved over time. The first generation consisted of small-scale models estimated by maximum likelihood and the Kalman filter (Engle and Watson, 1981; Mariano and Murasawa, 2003; Camacho and Perez-Quiros, 2010). These models were able to handle data irregularities, but were unable to utilize more than a few variables.

Forecasters and policymakers, however, monitor a large number of different time series (Bernanke and Boivin, 2003). Because the time span of most of the series is rather short — a problem of even bigger importance in economies that transformed to a market economy relatively recently — applying traditional models to a large number of variables would result in parameter proliferation and imprecise forecasts. Therefore, the second generation of factor models uses nonparametric principal component estimation of factors from large cross sections (Chamberlain and Rothschild, 1983; Forni and Reichlin, 1998; Forni et al., 2000; Stock and Watson, 2002a; Stock and Watson, 2002b). However, principal components cannot deal with ragged ends on their own.

The third generation of factor models combines the first and second generations: factors approximated by principal components are utilized within a state-space framework (Giannone et al., 2008; Rünstler et al., 2009; Bańbura and Rünstler, 2011). Thus, they constitute a model that can handle large data sets with data irregularities present in a real-time forecasting setting. The asymptotic properties of these models can be found in Doz et al. (2011).

Finally, the most recent papers use the expectation-maximization algorithm to obtain maximum likelihood estimates of large models that are able to deal with unbalanced datasets (Schumacher and Breitung, 2008; Bańbura and Modugno, 2010). On the whole, this approach consists of iterating between the two steps: estimating the parameters conditional on the factors, and estimating the factors conditional on the parameters from previous iterations. The asymptotic theory is provided in Doz et al. (2012).

An accessible survey of dynamic factor models can be found in Stock and Watson (2010), while Bai and Ng (2008) provide a more technical survey. Bańbura et al. (2010b), Bańbura et al. (2013) survey the application of factor models with a focus on nowcasting.

In our empirical exercise we will use the latest generation dynamic factor model estimated by the expectation-maximization algorithm. We begin by specifying the model for monthly variables:

$$x_t = \Lambda f_t + \varepsilon_t \tag{1}$$

$$f_t = A_1 f_{t-1} + \cdots + A_p f_{t-p} + u_t, \tag{2}$$

where x_t is a vector of monthly variables transformed into stationary ones, f_t is a vector of r (unobserved) common factors, and u_t is a vector

[2] Arnostova et al. (2011), in their replication of Rünstler et al. (2009), account for publication lags, but their analysis is based on a revised dataset.

[3] Note that not all of these papers use unrevised data, so the comparability should be interpreted with caution.